

RESEARCH ARTICLE

Open Access

Genomic analysis of the regulatory elements and links with intrinsic DNA structural properties in the shrunken genome of *Buchnera*

Lilia Brinza^{1,2,3}, Federica Calevro¹ and Hubert Charles^{1,2*}

Abstract

Background: *Buchnera aphidicola* is an obligate symbiotic bacterium, associated with most of the aphididae, whose genome has drastically shrunk during intracellular evolution. Gene regulation in *Buchnera* has been a matter of controversy in recent years as the combination of genomic information with the experimental results has been contradictory, refuting or arguing in favour of a functional and responsive transcription regulation in *Buchnera*. The goal of this study was to describe the gene transcription regulation capabilities of *Buchnera* based on the inventory of cis- and trans-regulators encoded in the genomes of five strains from different aphids (*Acyrtosiphon pisum*, *Schizaphis graminum*, *Baizongia pistacea*, *Cinara cedri* and *Cinara tujafilina*), as well as on the characterisation of some intrinsic structural properties of the DNA molecule in these bacteria.

Results: Interaction graph analysis shows that gene neighbourhoods are conserved between *E. coli* and *Buchnera* in structures called transcriptons, interactons and metabolons, indicating that selective pressures have acted on the evolution of transcriptional, protein-protein interaction and metabolic networks in *Buchnera*. The transcriptional regulatory network in *Buchnera* is composed of a few general DNA-topological regulators (Nucleoid Associated Proteins and topoisomerases), with the quasi-absence of any specific ones (except for multifunctional enzymes with a known gene expression regulatory role in *Escherichia coli*, such as AlaS, PepA and BoIA, and the uncharacterized hypothetical regulators YchA and YrbA). The relative positioning of regulatory genes along the chromosome of *Buchnera* seems to have conserved its ancestral state, despite the genome erosion. Sigma-70 promoters with canonical thermodynamic sequence profiles were detected upstream of about 94% of the CDS of *Buchnera* in the different aphids. Based on Stress-Induced Duplex Destabilization (SIDDD) measurements, unstable σ^{70} promoters were found specifically associated with the regulator and transporter genes.

Conclusions: This genomic analysis provides supporting evidence of a selection of functional regulatory structures and it has enabled us to propose hypotheses concerning possible links between these regulatory elements and the DNA-topology (i.e., supercoiling, curvature, flexibility and base-pair stability) in the regulation of gene expression in the shrunken genome of *Buchnera*.

Keywords: *Buchnera aphidicola*, Genome reduction, Transcription regulation, DNA-topology, Nucleoid associated proteins (NAPs)

* Correspondence: hubert.charles@insa-lyon.fr

¹UMR203 BF2I, Biologie Fonctionnelle Insectes et Interactions, INSA-Lyon, INRA, Université de Lyon, Villeurbanne, France

²BAMBOO, INRIA Rhône-Alpes, Montbonnot Saint-Martin, France

Full list of author information is available at the end of the article

Background

Buchnera aphidicola, associated with most of the aphids (Hemiptera: aphididae), is a fascinating bacterium, both because of its apparent minimalist physiology and because of its intermediate status between an autonomous cell and an intracellular organelle. Shaped by some 150–200 million years of intracellular evolution, its genome and regulatory system have evolved to fit the evolutionary constraints imposed by the symbiotic partnership [1,2]. This work is a comparative genomic analysis of cis- and trans-regulators encoded in the *Buchnera* genomes of five different aphid strains, *Acyrtosiphon pisum* (BAP), *Schizaphis graminum* (BSg), *Baizongia pistacea* (BBp), *Cinara cedri* (BCc) and *Cinara tujafilina* (BCT), combined with analyses of the intrinsic physical topological properties of their DNA molecules. The main objective was to decipher the regulatory mechanisms underlying gene regulation in these bacteria.

The *Buchnera* genomes from the five aphid species share certain properties: (1) a small size, from 416 kb for BCc to 641 kb for BAP [1,3–6]; (2) a low GC-content of about 25%; (3) a standard bacterial gene density of about 85% of coding DNA; (4) the conservation of most genes encoding enzymes from the biosynthesis of essential amino acids that *Buchnera* furnish to their hosts [7,8]. The differences between these five aphid species are related to the physiology of the symbiotic interactions that created specific evolutionary constraints, contributing to the differentiation of the *Buchnera* gene repertoires [9]. BCc offers an example of evolution with an extremely reduced genome, probably linked to the presence in their aphid host of the co-primary endosymbiont “*Candidatus Serratia symbiotica*” with which they show a strong dependency and partially share genes of several amino acid biosynthetic pathways [10,11].

Gene regulation in *Buchnera* has been a matter of controversy in recent years. Global transcriptomic analyses revealed a weak transcriptional response to various stresses applied on the host, such as heat shock [12] and single amino acid excess [13] in BSg, as well as aromatic and essential amino acid depletions in BAP [14]. However, stronger effects were observed when the transcriptional responses were compared between *Buchnera* populations from embryonic and maternal aphid compartments [15], somehow reflecting two different physiological growing states of *Buchnera*. Finally, following the kinetics of the response in BAP, a specific induction (repression) of the genes of the leucine biosynthetic pathway was observed after one day of treatment following a depletion (excess) of the leucine concentration in the aphid diet, although the transcriptional response was not significant after seven days of treatment [16].

In bacteria, two main interrelated processes govern gene transcription [17]. There is the “classical” mechanism, involving sigma and specific transcription factors binding DNA sequences located in the proximity of the transcrip-

tion initiation site of a gene and able to induce or repress transcription initiation by the RNA polymerase [18]. Then there is a more recently discovered mechanism based on the regulation of DNA topology controlled by Nucleoid Associated Proteins (NAP) and topoisomerases [19–23] and partially characterized by several physical parameters, such as DNA stability, curvature and supercoiling [24,25]. Both processes involve trans-regulatory factors (*i.e.*, proteins binding to DNA) interacting, with varying degrees of specificity, with the cis-regulatory elements (*i.e.*, DNA sequences). The regulatory mechanisms controlling transcription initiation are the most thoroughly described in free-living bacteria [18] and genome organisation suggests that they have also been conserved in *Buchnera* [26]. Other events responsible for transcriptional regulation, like termination, mRNA maturation and stability control, as well as translation regulation, are also important targets for gene expression regulation, possibly involving small RNAs [27,28]. These mechanisms have not been taken into account in this work. Finally, post-translational modification by reversible N^ε-lysine acetylation of transcription factors has been recently reported in bacteria and might directly affect gene expression [29]. However, it seems unlikely that this mechanism exists in *Buchnera* as the corresponding acetyltransferase (Pat or YfiQ) and the NAD-dependant deacetylase (CobB), described in *Salmonella enterica* and *Escherichia coli*, are lacking in *Buchnera*.

The bacterial chromosome is known to be associated with proteins which allow for a massive compaction and, at the same time, are able to dynamically regulate the DNA molecule, rendering rapidly accessible those DNA regions which need to be transcribed [30,31]. NAPs have been extensively described in *E. coli* [32,33]. They participate in the chromosome structuring and also in all the processes involving DNA transactions (replication, recombination and transcription). NAPs are basic, small molecular weight proteins and their relative abundance is dynamic and dependent on the cell physiology. For example, different bacterial growing phases are characterised by specific expression patterns of the different NAPs [34,35]. Although more than 12 NAPs have been described in *E. coli*, almost all the literature centres on only four of them: H-NS (Histone-like Nucleoid Structuring protein), HU (Heat Unstable nucleoid protein), IHF (Integration Host Factor) and FIS (Factor for Inversion Stimulation). Apart from NAPs, the maintenance of the chromosome supercoiling in bacteria is controlled by topoisomerases that either relax the negative supercoils (type I topoisomerase) or serve to introduce them (ATP consuming type II topoisomerase), hence linking the energetic metabolism of the cell with the DNA topology [36].

Negative supercoiling is essential for chromosome compaction and for the survival of the bacteria [21,37]. Local supercoiling variations in the DNA modulate the

polymerase affinity for promoters. Hence, DNA supercoiling is often considered as a true transcriptional regulator that is sensitive to the environmental conditions [19,23,35,38] and that uses the ATP/ADP ratio governing gyrase activity as a sensor of the energetic level of the cell [36,39]. Curvature, flexibility and stability, contrary to supercoiling, are properties which are highly correlated with the primary sequence of the DNA molecule, although it has been suggested that nucleoid proteins might also influence these parameters [40]. Curvature and flexibility (estimated in this work by the base-pair propeller twisting) are essential for the initiation of transcription since the promoter affinity for polymerase, and for the associated transcription factors, is sensitive to the topology of the DNA molecule [24]. The double-strand stability of the DNA molecule is also very important, particularly in the promoter region, for transcription initiation. In this study, we have estimated the base stacking energy, which is a direct measure of the base pair affinity within the DNA molecule [41], and the Stress-Induced DNA Duplex Destabilization (SIDD) [42] to assess the stability of the double strand DNA molecule in *Buchnera*.

The aim of this work is to give an initial description of the structure and of the evolution of the gene regulatory network in *Buchnera* using a genomic comparative analysis. Regulatory networks are known to evolve quickly and both the DNA-binding domains of transcription factors and their target sequence sets are highly dynamic (i.e., orthologous regulators are often regulating non-orthologous targets), making comparative studies difficult [43,44]. The *Buchnera* model is interesting in this respect, first because the bacteria evolved for millions of years sequestered within aphids, preventing any contact with other bacterial populations and, hence, any horizontal gene transfer and, secondly, because *Buchnera* were almost uniformly constrained by the intracellular conditions (which relaxes the selection of some genes that become superfluous) and by the physiological requirements imposed by their symbiotic association with aphids (mostly concerning the biosynthesis of essential metabolites, such as amino acids). Thus, after analysing the selective constraints exerted on the regulatory genes, we performed a systematic characterization of the cis- and trans-regulatory elements predicted in the *Buchnera* genomes from the five sequenced strains. These analyses were then coupled with the characterisation of some intrinsic topological properties of the *Buchnera* DNA chromosome, which allowed us to formulate certain hypotheses regarding their possible involvement in gene transcription regulation in these bacteria.

Methods

Buchnera strains and GenBank sequence accession numbers

Four *B. aphidicola* genomes (*BAp*, *BSg*, *BBp*, *BCc*) from the following four aphids: *Acyrtosiphon pisum*, BA000003

[1]; *Schizaphis graminum*, AE013218 [3]; *Baizongia pistaciae*, AE016826 [4]; *Cinara cedri*, CP000263 [5], were used in this work. A fifth genome sequence was published recently from the aphid *Cinara tujafilina* (CP001817 [6]), very closely related to *BCc*. As *BCt* and *BCc* show very similar genomic repertoires and properties, *BCt* was not always included in the analyses of this study. The annotations of transcription units (operons) in *BAp* are those described in [26].

In this work, the *E. coli* genome was used as the reference ancestral state of the *Buchnera* lineage, disregarding the evolution in the branch of *E. coli*. The analyses were performed on the genomic sequence of *Escherichia coli* str. K-12 substr. MG1655 (U00096) [45,46] and information concerning the gene regulatory network was taken from the RegulonDB 6.2 database [47].

Inventory of the transcription factors encoded in *Buchnera* genomes

The inventory of the transcription factors encoded in the *Buchnera* genomes was primarily carried out using the close proximity of *Buchnera* to *E. coli* and the one-to-one orthology relationship for almost all of the genes of *Buchnera*, as referred to in BuchneraBase [48].

We completed this inventory by searching in the *Buchnera* protein set for all Pfam domains annotated as regions with putative regulatory functions [49]. As helix-turn-helix (HTH) domains are by far the most common protein DNA-binding domains in bacteria [50,51], a systematic search for these structural domains was also carried out for the complete set of proteins in *Buchnera*, using the HTH software [52,53]. Non-HTH motif predictions such as zinc fingers, helix-loop-helix, beta-sheet antiparallel or RNA binding domains are available in the generalist DNA Binding Domain Database [54]. Indeed, three non-HTH predictions are available for *Buchnera*: CspE and CspC each with a cold-shock domain and DksA with a zinc finger domain. These proteins were already detected by their orthologous annotations in *E. coli*.

Annotation of the cis-regulatory elements in *Buchnera*

Promoters and transcription starts were predicted, in the four *Buchnera* strains, using the software BPROM[®] (<http://linux1.softberry.com/berry.phtml>) and MacVector (MacVector, Cary, NC, USA) for σ^{70} and σ^{32} promoters, respectively. BPROM was calibrated for *E. coli* with a specificity of about 80%. In this work, predictions were made using the *E. coli* parameters in the 500 bp regions located upstream from all the *Buchnera* coding DNA sequences (CDS). The motif detection function of MacVector was calibrated using the two consensus σ^{32} -boxes (CTTGAAAA and CCCCTNT), separated by 11–15 bp [55], with a tolerance of 50% similarity, according to previous work performed in *BSg* [12].

The transcription factor binding sites (TFBS), known in *E. coli* for the different NAPs, were searched for in the *Buchnera* genomic sequences using the *words.pos* and *gregexpr* R-functions from the *SeqinR* [56] and *base* libraries, respectively. The following consensus motifs were searched: GNTYAWWWWWTRANC for FIS [57], WATCAANNNTTR for IHF [58] and TCGWTWAAWW for H-NS [59].

Statistical analysis and gene ontology annotations

All the statistical analyses were performed using the R software (<http://www.r-project.org>). The Gene Ontology (GO) annotations were extracted from the UniprotKB-GOA database [60]. Annotations were compared at the same level (i.e., level 3 or 4) to avoid redundancy-bias linked to the possible non-homogenous depth of the different branches of the GO hierarchy.

Interaction-graph analysis and genome rearrangements

The C3P software, developed by Boyer *et al.* [61], is a graph-theoretical approach extracting common connected components between two or more graphs for exploring gene neighbourhoods in a genomic and functional context. If we consider two genes, *x* and *y*, located in close proximity on the genome, the main principle of this method is that if *x* and *y* are co-regulated (i.e., neighbours in the transcription regulatory network), or if the proteins *X* and *Y* are interacting directly (i.e., neighbours in the protein-protein interaction network), or are catalyzing successive steps in a metabolic pathway (i.e., neighbours in the metabolic network), the relative positioning of the genes *x* and *y* should be preserved during genome evolution if selective pressures are acting on the transcriptional, protein-protein interaction or metabolic networks, respectively. In the present study, we have applied this approach to extract the common connected components between the *Buchnera* genome, on the one hand, and the transcription network, protein interaction network and metabolic network on the other hand. Two genes are considered as neighbours if they are separated by a maximum of one gene in the molecular interaction networks and by five genes on the genome. A set of neighbour genes (or of proteins encoded by a set of neighbour genes) is called: (1) a synton in the genome; (2) a transcripton (this term was not defined in the original work of Boyer *et al.* [61]) in the transcriptional regulatory network; (3) an interacton in the protein-protein interaction network; or (4) a metabolon in the metabolic network. The transcriptional network, together with the protein-protein interaction and the metabolic networks in *Buchnera*, were inferred from *E. coli* by direct orthology. In order to test the significance of the structures described in *BAP* (i.e., to establish whether they could have been observed by chance), we developed a procedure to simulate random transcriptons

(r-transcriptons). For that, *X* genes were randomly selected, within a limited span of 5 genes in the *E. coli* genome, to generate an r-transcripton. The size (*X*) of each r-transcripton was defined by sampling the sizes of the true transcriptons of *E. coli*. The r-transcriptons were retained only if they shared at least two orthologous genes with *BAP*. Each simulation was ended either when 38 r-transcriptons, sharing at least two orthologues with *BAP*, were generated or when all the genes of *E. coli* had been used since gene sampling was performed without replacement. Since a number of r-transcriptons equal to, or greater than 37 were found in only 9 out of the 1000 simulations, a *p*-value of $9 \cdot 10^{-3}$ was estimated (see Results and discussion section and Table 1 for the justification of the thresholds).

We also analysed the evolutionary history of each transcription unit (TU) making up the *E. coli* transcriptons and conserved in *BAP*. For that, several types of TUs were defined: identical TUs are *BAP* TUs with exact orthologous replicates in *E. coli* whereas similar TUs are *BAP* TUs that underwent gene deletions in *BAP* (or more rarely gene insertions in *E. coli*). Split, merged and reorganized TUs are *BAP* TUs that were recomposed from different “ancestral” TUs during *Buchnera* evolution. More details and schemes are given in [26].

Physical properties of the DNA molecule of *Buchnera*

Four physical sequence-dependent properties of the DNA molecule were computed using the GeneWiz software [25] for the four sequenced *Buchnera* strains: stress-induced duplex destabilization - SIDD [42], curvature [62], base stacking energy [41] and base-pair propeller-twisting [63]. The calculation of these different parameters with GeneWiz leads to a single value for each base of the DNA sequence. Non-overlapping sliding windows of different sizes were then defined to calculate the smooth parameter-distributions at different scales (e.g., 300 bp for the whole genome sequence analysis and 20 bp for the promoter region analysis).

For the whole genome comparison between *Buchnera* and *E. coli*, two null models of base composition were constructed. The first one (global null model) is a uniform permutation of all the bases of the genome (i.e., preserving only the global GC-content on the genome scale) whilst the second one (local null model) is a uniform base permutation applied to each coding and non coding region (i.e. preserving the local GC-content, as well as the coding versus non-coding GC content).

Results and discussion

Selective constraints acting on the regulatory structures and genome functional analysis in *Buchnera*

In order to determine whether protein interactions (across protein, metabolic or transcription regulatory networks) exerted a selective pressure on the conservation of some

Table 1 Fate, in *BAp*, of “ancestral” orthologous interactons, regulons and metabolons found in *E. coli*

	# of <i>E. coli</i> CCC ^a	# of <i>E. coli</i> CCC with 2 genes conserved (at least) in <i>BAp</i>	# of <i>E. coli</i> CCC conserved as CCC in <i>BAp</i>	# of <i>BAp</i> CCC not found in <i>E. coli</i>
Transcriptons	239	38	37	12
Interactons	39	10	9	0
Metabolons	107	23	23	0

^a: Number (#) of Common Connected Components.

genomic regions in *BAp*, we used the C3P program [61] to calculate the number of interactons, metabolons and transcriptons conserved between *E. coli* and *BAp* (see Methods section for the definitions). Essentially the idea is that if selective constraints have an effect on the three latter structures, and if the genes are close neighbours on the *E. coli* genome (supposed, here, to reflect the ancestral state), we expect to see conservation of the corresponding orthologous genes in the equivalent neighbourhood in *BAp*. Indeed, Table 1 shows that 37/38, 9/10 and 23/23 of the *E. coli* transcriptons, interactons and metabolons, respectively, found in *BAp* are conserved as common connected components (so the proximity between genes was conserved between *E. coli* and *Buchnera* for these particular genes). These results are significant (*i.e.*, not observed by chance), as revealed by a re-sampling test giving a p-value of 0.009 for the *BAp* transcriptons (see Methods section). The procedure was not applied to the metabolons and interactons because they are bigger structures, with a lower probability of being observed randomly as conserved associated structures.

Moreover, we have shown that, in addition to the 37 “ancestral” transcriptons, 12 supplementary ones have been produced in *BAp*, by genomic rearrangements, since its divergence with *E. coli*. We also analysed the evolutionary history of the TUs composing the transcriptons, found in *BAp*, following on from our previous work on TUs in *Buchnera* [26]. Among the 55 TUs composing the 38 transcriptons found in both *BAp* and *E. coli*, 16 are monocistronic, 13 are identical or similar TUs (totally or partially conserved from gene deletions) and 26 are TUs that were reorganised during genome evolution, *i.e.*, formed from a fusion or split from different ancestral TUs (see Methods section for definition of the different TU rearrangements). These results reveal that these two bacterial lineages conserved not only some single operonic structures but also some bigger syntenic fragments associating several TUs. Moreover, genomic rearrangements occurring in the *BAp* lineage during genome shrinkage seem to have clustered some co-regulated genes within the same neighbourhood (*i.e.*, new transcriptons).

During the process of genome shrinkage in the *Buchnera* lineages, which followed the symbiotic association with

aphids, gene losses occurred preferentially within functional classes that escaped from selective pressure due to the new intracellular environment. A systematic search for under- and over-represented functional classes of genes was performed in *BAp*, relative to those found in *E. coli* (Additional file 1A), with a specific focus on gene expression regulation. Hence, as previously mentioned in the primary annotation [1], gene proportions within the GO classes of *transporter activity*, *developmental processes*, *response to stimulus*, *localization* and *biological regulation* are significantly lower in *Buchnera*, compared to *E. coli*, whereas proportions of many classes corresponding to central core metabolism (*e.g.* *catalytic activity*, *metabolic processes* and *cellular processes*) are significantly higher.

Moreover, we demonstrate here that the genes preferentially conserved in *BAp* are mostly the multifunctional ones. Indeed, conserved genes in *BAp* have a significantly higher association with multiple GO annotations since the distribution of the number of GO terms associated with *BAp* genes shows more genes with 2 to 8 GO terms, when compared to *E. coli* (Additional file 1B, Wilcoxon rank test p-value = $2 \cdot 10^{-16}$). These multifunctional genes are mostly metabolic genes (*i.e.*, associated with the GO term GO:0008152). Indeed, when the analysis is repeated, after removing those genes associated with the corresponding GO term, the distributions remain almost equivalent (Additional file 1C, Wilcoxon rank test p-value = 0.07). This result is important in the context of genome shrinking, demonstrating that genes encoding multifunctional proteins may be favoured in small genomes.

Buchnera have lost a number of genes from the catabolic pathways since most of the end products, synthesized by *Buchnera*, are exported to the host (only 9% of the *E. coli* catabolic pathways are present in *BAp*). Salvage pathways have been interwoven between the two symbiotic partners during symbiotic evolution [8,64]. It is important to note that catabolic genes are generally more highly regulated in *E. coli*, as compared to anabolic genes [65]. Indeed, 83% of the catabolic genes are regulated in *E. coli*, versus only 40% for the anabolic genes. Hence, the loss of catabolic genes partly explains the decay of the gene expression regulatory system in *BAp*.

Inventory of the trans-regulatory elements in *Buchnera* genomes

A systematic search was performed for the complete set of proteins of the four *Buchnera* strains (*BAp*, *BSg*, *BBp* and *BCc*) using one-to-one orthology with annotated genes of *E. coli*, scanning for annotated Pfam domains and HTH domains (see Methods section). Nineteen proteins were detected and are presented in Table 2. The corresponding orthologous genes were searched for within the newly sequenced *BCt* genome and are

Table 2 Inventory of regulatory proteins detected in *Buchnera*

Gene name	Regulator type	Identity ^a	BAP	BSg	BBp	BCc	BCt ^b	Prot ^c
<i>rpoD</i>	σ factors	82.11	+	+	+	+	+	+ (236)
<i>rpoH</i>		72.18	+	+	+	+	+	–
<i>alaS</i>	"Specific"	50.11	+	+	+	+	+	+ (137)
<i>bolA</i>		27.62	+	+	+	+	+	–
<i>pepA</i>		52.09	+	+	+	–	+	+ (301)
<i>metR</i>	Bifunctional	–	pseudo	+	–	–	–	–
<i>dksA</i>		70.89	+	+	+	–	+	+ (218)
<i>cspC</i>		91.30	+	–	–	–	–	+ (229)
<i>cspE</i>		94.20	+	+	+	+	+	+ (243)
<i>csrA</i>		86.89	+	+	+	+	+	–
<i>ychA</i>	Hypothetical	57.62	+	+	+	+	+	+ (215)
<i>yrbA</i>		41.67	+	+	+	+	+	+ (348)
<i>fis</i>	NAPs and topoisomerases	65.31	+	+	–	+	–	+ (366)
<i>hns</i>		60.58	+	–	–	–	–	–
<i>hupA</i>		66.30	+	+	+	–	–	+ (307)
<i>himA</i>		59.80	+	+	–	+	+	+ (321)
<i>himD</i>		68.08	+	+	–	+	+	+ (397)
<i>ybaB</i>		81.65	+	+	+	+	+	+ (283)
<i>topA</i>		55.43	+	+	–	–	–	–
<i>gyrA</i>		63.58	+	+	+	+	+	+ (102)
<i>gyrB</i>		62.11	+	+	+	+	+	+ (124)

^a: protein sequence identity (%) with the *E. coli* orthologous gene; ^b: genome sequence of *BCt* was not analysed systematically, we have only reported here the presence/absence of orthologous genes; ^c: proteomic data were compiled from [66], +/– (rank) indicates the presence/absence of the protein in partially purified samples of *BAP* (rank of the quantitative detection out of 400 proteins).

presented in the same table. Although no new regulator was discovered, this is the first time that such a global and comparative analysis across the *Buchnera* strains has been published.

Sigma factors

Among the six sigma factors that were predicted in the last common ancestor of *E. coli* and *Buchnera* [67,68], only two have been conserved in the five *Buchnera* strains analysed in this work: σ^{70} , encoded by the *rpoD* gene, which is the constitutive bacterial sigma factor, and σ^{32} , encoded by the *rpoH* gene, which is the factor responsible for heat shock regulon transcriptional control. These two proteins show a high sequence identity with those of *E. coli*, and the RpoD protein was detected by proteomics analysis in a partially purified *BAP* sample, whereas RpoH (probably repressed in non-stressed conditions) was not ([66] in Table 2). Hence, the σ^{24} , σ^{28} , σ^{38} and σ^{54} regulons have all been lost during the process of genome reduction in the *Buchnera* lineage, and this loss was probably a long time ago as it occurred before the divergence that gave rise to the evolution of the five *Buchnera* strains analysed here.

Transcription factors

Four transcription factors are described as "specific" in our work as they are associated with one, or only a very few, target genes in *E. coli*. Three of them (*AlaS*, *PepA* and *BolA*) are present in the genomes of the all *Buchnera* strains we analysed, whereas most of their known putative targets in *E. coli* are not conserved. Only *AlaS* and *PepA* were detected by proteomics analyses performed in the *BAP* strain [66]. *AlaS*, encoding alanyl-tRNA synthetase, acts as an autorepressor of the *alaS* gene, sensitive to the concentration of alanine in bacterial cells [69]. In *E. coli*, the gene *pepA* encodes the multifunctional DNA-binding enzyme *PepA*, which is an aminopeptidase also acting as a transcription factor regulating the *carAB* operon and, thus, it is involved in the metabolism of arginine and proline. The fixation domain to the DNA of *PepA* has been widely studied as it is atypical and does not show classical DNA-binding motifs [70].

A similar example of the conservation, in *Buchnera*, of such multifunctional enzymes with transcriptional regulatory properties is provided by the reductase *BolA* known in *E. coli* for forming iron-sulfur (FeS) clusters with glutaredoxin [71,72]. The expression of *BolA* was originally described as being exclusively associated with

the stationary phase in *E. coli*, but BolA is also implicated in the response to a broad range of stress conditions (heat, osmotic, oxidative, acidic and nutritional stresses) [73]. In addition, this protein is described as a morphogene, containing a putative HTH-domain potentially involved in the regulation of genes responsible for the cellular morphology changes in stress conditions [74-76]. Until now, BolA has been shown to be able to regulate only four genes, at a transcriptional level, in *E. coli*: the *dacA*, *dacC* and *ampC* genes (involved in penicillin resistance) and the *mreB* gene (involved in rod-shape maintenance) [71,72]. None of these genes have been conserved in the *Buchnera* genomes. Moreover, in *Buchnera*, the HTH DNA-binding domain was not detected with a statistically significant score. Several studies have recently highlighted the transcriptional regulatory role of enzymes in bacteria [77-80]. In *Buchnera*, the absence of two-component sensors might be compensated for by such multifunctional enzymatic systems that combine catalytic and regulatory properties and are directly sensitive to substrate availability. This hypothesis is consistent with the observation that *Buchnera* genomes have selected for the conservation of multifunctional genes (Additional file 1) but it is not, as yet, known whether AlaS, PepA and BolA have acquired a broad regulatory role and a large spectrum of targets, nor whether other enzymes might have been recruited for a similar regulatory function in *Buchnera*.

Finally, the regulator *metR*, involved in the transcriptional regulation of the methionine biosynthetic enzymes, is conserved only in *BSg* whereas it is present as a pseudogene in *BAp* and absent in *BBp*, *BCc* and *BCt*. Moran et al. [13] have proposed an evolutionary scenario hypothesizing that the conservation of methionine regulation in *BSg* could be linked to the high variability of the cysteine concentration, supplied as homocysteine, in the *BSg* diet, compared to the other *Buchnera* strains.

Four transcription factors are referred to here as “bifunctional” (CspC, CspE, CsrA and DksA) because of their putative capability to bind both DNA and RNA molecules, although they do not function as classical transcription factors in *E. coli*. CspC and CspE are members of the cold-shock protein family of *E. coli*, which are also RNA chaperones thought to facilitate translation at low temperatures by destabilizing the mRNA secondary structures. Bae et al. [81] have revealed that these proteins also act as transcription anti-terminators, interacting with the Rho-independent termination mechanism. In addition, these proteins are known to play a role in the regulation of DNA-topology, stabilising chromosome compaction and, hence, regulating gene expression [82]. CspC is found only in *BAp*, whereas CspE is conserved in the genomes of the five sequenced *Buchnera* strains. CsrA, conserved in the five *Buchnera* strains, has been identified in *E. coli* as a post-transcriptional

regulator able to bind to the 5' untranslated leaders of some mRNAs by competing with the interaction of the 30S ribosomal subunits, thus inhibiting their translation [83]. CsrA may also be involved in indirect gene expression activation by an, as yet, unknown mechanism [84,85]. In *E. coli*, CsrA is involved in many cellular processes but it was originally discovered as a global regulator of carbon source metabolism. It is a general activator of glycolysis (which *Buchnera* can do) and a repressor of glycogenesis (which *Buchnera* cannot do). CsrA is antagonized, in *E. coli*, by the sRNAs CsrB and CsrC that seem to be absent in *Buchnera*. Edwards et al. [83] have recently proposed a more general regulatory role for CsrA, describing 721 putative target genes in *E. coli*, and they have revealed a link with the stringent response involving the DksA protein. DksA is known, in *E. coli*, for its ability to regulate transcription elongation. The protein is able to bind with the RNA polymerase secondary channel, to interact with the alarmone ppGpp, and to destabilize the transcription complexes bound at the discriminator sites of the promoters, hence allowing elongation to start [86-89]. Although the ppGpp alarmone can no longer be produced in *Buchnera* (due to the absence of *relA* and *spoT* genes), DksA could have conserved its regulatory roles in this bacterium. We hypothesize that DksA might destabilize the σ^{70} factor associated with the AT-rich discriminators of most genes, hence promoting transcription elongation in these AT-base enriched genomes. In *BAp*, *BSg*, *BBp* and *BCt*, DksA might also interact with the elongation factor GreA (absent in *BCc*). GreB (a paralogous elongation factor found in *E. coli*) is absent in the genomes of the five *Buchnera* strains analysed in this work. DksA and GreA show similar binding properties with the secondary channel of the RNA polymerase and show functional redundancy in *E. coli* but have also opposite regulatory effects for some genes [90]. GreA and DksA might have been conserved together in most *Buchnera* strains for this opposite effect or also for the chaperone properties of the GreA protein [91] but this remains yet speculative. It is to note that *BCc* lost both DksA and GreA proteins.

Two proteins, YchA and YrbA, referred to here as “hypothetical regulators”, are conserved in the five *Buchnera* strains. YrbA is an homologous protein of BolA and was very recently renamed IbaG (induced by acid gene) in *E. coli* [92]. IbaG doesn't share the morphogene properties of BolA and its enzymatic activity was not analysed but it seems to act as a transcriptional regulator to protect the cell against stress. YchA is annotated in *E. coli* as putative transcriptional regulators. Nothing is known about the function of these two proteins in *Buchnera*.

In the classes of “bifunctional” and “hypothetical” regulators, most of the genes (except *cspC* that is found only

in *BAp*) are conserved in the five *Buchnera* strains. Expression data, published by Vinuelas *et al.* [93], revealed that *dksA*, *cspC*, *cspE* and *csrA* are all highly expressed and highly conserved in *BAp*, *ychA* is highly conserved with a low expression level, and *yrbA* is evolving quickly and is highly expressed. Since Poliakov *et al.* [66] detected the proteins DksA, CspC, CspE, YchA and YrbA by proteomics analysis of *BAp* samples, these six genes represent interesting candidates for future experimental studies.

Nucleoid associated proteins and topoisomerases

Seven regulators of DNA-topology are conserved in the *Buchnera* genomes whereas most of the specific transcriptional regulators have been lost. Among them, five are classified as NAPs: FIS, H-NS, HU (encoded by *hupA*), IHF (encoded by *himA* and *himD*) and YbaB, and the other two are the topoisomerases TopI and the DNA Gyrase (encoded by *gyrA* and *gyrB*).

The identity percentages between the *E. coli* and *Buchnera* protein sequences are high for the NAPs and their functionally important sites (when known) are all well conserved, as reported in Table 2 and illustrated in Additional file 2. One exception is the FIS protein that is composed of two domains. The N-terminal domain, required for phage integration and recombination activity, has accumulated amino acid substitutions in *Buchnera* genomes lacking recombination capability. On the other hand, the C-terminal DNA-binding domain of the protein is highly conserved (Additional file 2).

With the exception of YbaB (function unclear, as stated by Dillon and Dorman [33]), the NAPs are not equally conserved in the genomes of the five *Buchnera* strains and not all have been detected in *BAp* by proteomic analysis [66] (Table 2). We have observed that when HU is lacking as it is in *BCc*, IHF is present, and vice versa as observed in *BBp*. *BBp* shows the smallest set of NAPs (with only YbaB and HU). Hence, the conservation of a minimal set of NAPs is consistent with a pleiotropic function of the two paralogous proteins, HU and IHF, which can compensate each other in *E. coli*, and also with the non-viability of deletion mutants lacking HU, IHF and H-NS [94].

Analysis of the syntenicity between the different *Buchnera* genomes reveals that the loss of each NAP is a specific gene deletion, as neighbouring genes on both sides are always well conserved (data not shown). Hence, these observations are consistent with the loss, in the different strains, of some selective pressure on gene compaction and global repression. Such differences of selection pressure between the *Buchnera* strains have also been reported for the transport function, and they have been correlated with the differential success of their aphid hosts [9]. *A. pisum* and *S. graminum* are modern, cosmopolitan, oligophagous and successful aphids and their *Buchnera* have

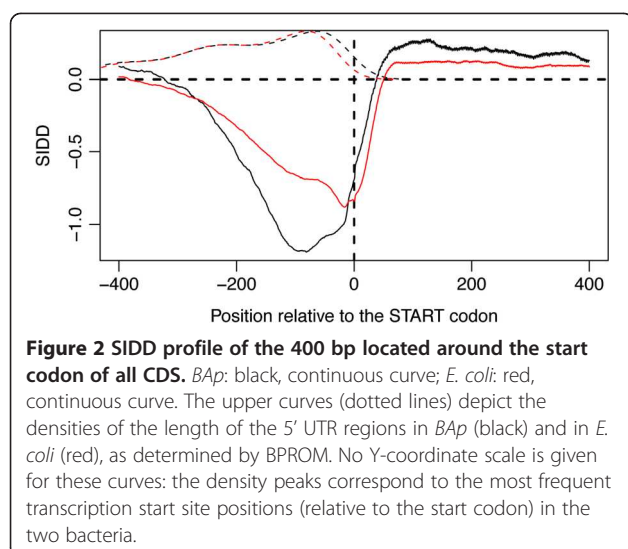
retained the largest set of NAPs, whereas *B. pistaciae*, *C. cedri* and *C. tujaefilina* are more primitive aphids, with smaller geographical distribution and slower growing rates, and their *Buchnera* have retained minimal sets of NAPs, possibly because these bacteria are thought to face less diverse physiological conditions.

NAPs are global regulators of DNA topology but they can also act as true transcription factors capable of interacting with the promoter regions of some target genes. Transcriptomic analyses revealed that mutants deleted for NAPs, or overexpressing them, modify the gene expression regulation of numerous genes (e.g., 819 and 610 for Δ FIS and Δ HNS mutants, respectively, in *E. coli* [19]). However, among these genes, subgroups of more specific targets are found in the RegulonDB database [47]. We have analysed, in *BAp*, the distribution of the conserved specific target genes (using *E. coli* as the reference ancestral model) for the five NAPs (Additional file 3). Since a neutral model of random gene deletion would predict equivalent proportions of conserved target/non-target genes for all the NAPs, no specific selection for gene conservation was found in the *Buchnera* strains analysed (Additional file 3).

Buchnera possess the minimal set of topoisomerases required to control chromosome supercoiling: one topoisomerase I, which removes negative supercoils without consuming ATP, and one gyrase (ATP-dependent) necessary for their introduction. In *E. coli*, two supplementary topoisomerases are involved in the decatenation process, allowing for chromosome separation at the end of replication. The absence of these two topoisomerases in *Buchnera* may be linked with the high ploidy of these bacteria, as observed by Komaki and Ishikawa [95] in *BAp*. More surprising is the absence of topoisomerase I in *BBp* and *BCc*. Gil *et al.* [96] have proposed that the gyrase in *Buchnera* might have acquired a broader function and has become capable of removing or introducing negative supercoils, as suggested previously by Drolet *et al.* [97] in *E. coli*.

Recent studies by Sobetzko *et al.* [39] show that the positioning of the different NAPs (as well as that of several other regulatory and metabolic genes) is not random on the chromosome but, instead, it is specifically associated with the DNA macrodomains that are differentially sensitive to the states of DNA-relaxation. This is observed across the range of bacterial diversity and provides an ancestral mechanistic insight into how the chromosome organisation “encodes” a spatiotemporal program to globally regulate gene expression. Although indicated by a previous genomic analysis [93], the existence of macrodomains has not been demonstrated in the *Buchnera* genomes. However, such ancestral genome organization (i.e., the relative positioning of NAPs along the chromosome) seems to be conserved in *BAp*, despite its genome erosion (Figure 1).

Figure 1 Spatial organisation of the *BAp* chromosome compared with that of *E. coli*. First line bars: genes involved in aerobic/anaerobic metabolism (dark blue), DNA replication (orange), *rrn* genes (upper red), and transition phase (brown). Second line bars: selected genes involved in the control of DNA topology (green). Third line bars: selected genes encoding NAPs (light blue). Fourth line bars: selected genes involved in modulating RNAP activity, including σ factor-utilization regulators (light green), secondary channel-binding proteins (pink), termination/elongation factors (red), and RNAP subunits (black). Grey dotted lines link pairs of orthologous genes between *BAp* and *E. coli*. Selected genes are those described by Sobetko *et al.* [39].



found. Nevertheless, it is significant that, in *Bap*, among the 45 conserved genes from the σ^{32} regulon of *E. coli*, only 19 (42%) had retained a recognizable σ^{32} promoter. Moreover, when comparing the putative σ^{32} regulons of these four *Buchnera* strains, they appeared to be very divergent, revealing that predicted σ^{32} promoters are not tractable in *Buchnera* without any experimental data (Additional file 6). These results are quite important as regards the possible role of σ^{32} in the AT-rich genomes of insect endosymbionts, characterised by mild transcriptional changes in response to heat shock [98,99]. It has been proposed that, in symbiotic genomes, this transcriptional regulator might control currently unknown stress signals, or it may have been transformed into an alternative vegetative σ factor in order to replace the lost σ factors (σ^{24} , σ^{28} , σ^{38} and σ^{54}). Indeed, the very high number of σ^{32} promoters found in the *Buchnera* genomes, coupled with the inconsistency of the predicted corresponding regulons, is likely to reflect a high rate of false positives rather than the acquisition of a more generalist role for this σ factor.

A systematic search of TFBS was also performed for the three NAPs for which a known TFBS has been described (see Methods section). However, in *Buchnera*, all the predicted TFBS are under-represented as regards what would be randomly expected, so it was not possible to extract significant TFBS even for these proteins (data not shown).

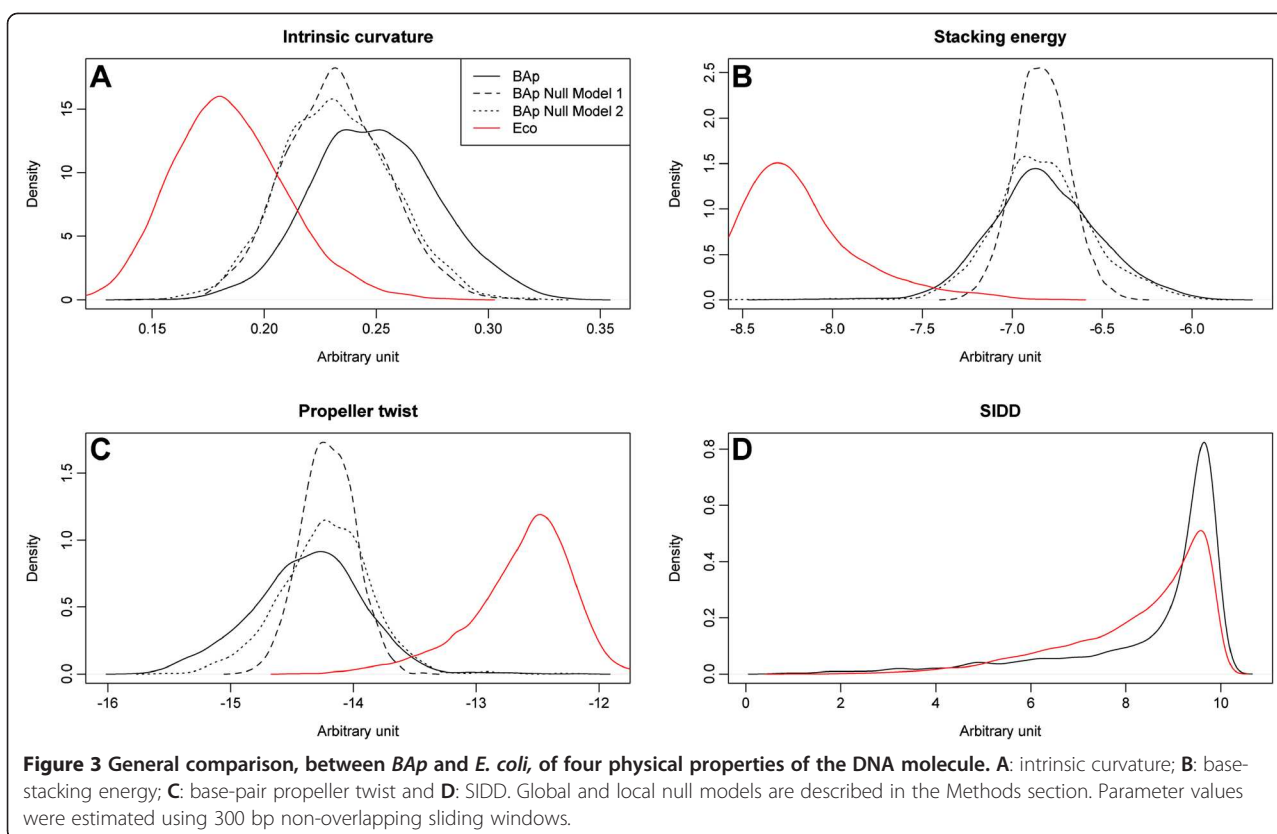
Intrinsic structural properties of the DNA molecule of *Buchnera*

The role of DNA negative supercoiling in the regulation of transcription activity in bacteria is now well established [19]. Here, we analysed the sequence dependent features determining some of the physical properties of the DNA molecule of *Buchnera*, namely intrinsic curvature and SIDD. These parameters enabled us to highlight the most

favourable promoter regions for transcription initiation, together with base stacking energy and base-pair propeller-twisting relevant to the flexibility and stability of the DNA. Physical properties of the DNA molecule of *Bap* were compared to those of *E. coli* using the global and local null models of base-composition as references (see Methods section). The results are presented in Figure 3. As a direct effect of its AT-richness, the DNA molecule of *Bap* is much more curved, flexible and shows less base-pair stability compared to that of *E. coli*. The curvature and the flexibility provided by the propeller twist are more pronounced in *Bap* than in the global and local null models.

The SIDD is less closely correlated to the AT-composition and it does not allow for discrimination between *E. coli* and *Bap*. In order to detect regions more prone to harbour functional promoters, we analysed the SIDD scores within the intergenic regions of *Bap*. Indeed, divergent intergenic regions (containing two 5' UTR sequences) encode more promoters than convergent ones (containing two 3' UTR sequences). Hence, the SIDD scores serve to separate the clearly divergent (less stable) and the convergent (more stable) intergenic regions. Interestingly, tandem regions (containing one 3'UTR and one 5'UTR), that could harbour either a TU external promoter, a TU internal promoter or even no promoter at all, reveal a bimodal distribution separating out the stable and the unstable regions (Figure 4A). We analysed, in more detail, the SIDD distribution within the tandem regions (Figure 4B) and we found that the intra-TU promoter regions were the most stable, whereas the inter-TU promoter regions still showed a bimodal distribution of stable and unstable regions. Referring to our previous work on TUs in *Bap* [26], it is probable that stable promoter regions correspond to false predicted inter-TU regions in this strain since long TUs, associating non-functionally related genes, seem to be over-abundant in *Bap*.

We analysed the correlation between the SIDD values of each gene promoter region (calculated in a window of 150 bp, located upstream of the CDS start) and the corresponding functional gene classes in *Bap*, using the *multiFun* classification [100]. Two classes were found to be significantly correlated with unstable promoter regions (Wilcoxon rank test): the transporter class (p-value = 0.03) and the regulator class (p-value = $2 \cdot 10^{-4}$). Significant gene lists are given in the Additional file 7. The GC-content of these intergenic regions are not significantly different from the overall intergenic GC-content of *Bap*, hence this correlation cannot be explained only by a strand bias or a local GC-bias effect (data not shown). The association between unstable promoters and genes involved in transcription or transport was described for 38 out of 43 free living bacteria analysed by Wang and Beham [40], whereas the authors pointed



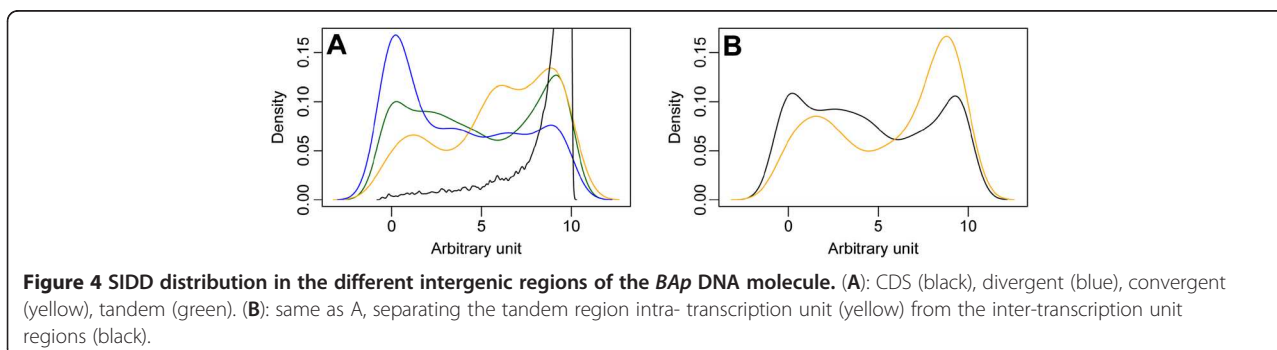
out that this regulatory mechanism was lost in a group of obligate parasitic bacteria (for 14 out of the 18 analysed) including *Chlamydia* and *Mycoplasma* species.

We also looked for associations between unstable promoters and functional gene properties in a data set separating the sensitive from the insensitive genes with respect to supercoiling variations in *E. coli* [19,23], as well as in several expression data sets of *BAp* involving stimulations of the essential amino acid metabolism [14,15,93]. However, no significant correlation was found (data not shown) probably because sensitive genes are different between *BAp* and *E. coli* and because the expression data were average for the total *Buchnera* populations within aphid tissues (i.e., maternal and embryonic populations with different nutritional requirements and stress sensibility).

Conclusions

Towards a putative model of topological regulation of the gene expression in *Buchnera*

The inventory of the regulatory elements of *Buchnera* reveals the very low diversity of specific regulators and the conservation of several NAPs and topoisomerases. These results are also consistent with the attenuated expression profiles observed in *BAp* and *BSg* microarray experiments [12-14] since NAPs induce continuous changes in the gene transcription rates, whereas local transcription factors induce discrete changes (i.e., On/Off transcription rates) [19,101]. DNA-superhelical density has not, as yet, been measured in *Buchnera* and changes in the supercoiling induced by variations of the intracellular environment of their bacteriocytes remains speculative. Hence,



the question of the functionality of DNA-topological regulators and of their links with gene expression regulation is still open.

Several evidences in favour of a selection of regulatory structures in *Buchnera* despite genome erosion are provided by this genomic analysis: (1) the conservation and the creation of new transcription structures during genome shrinkage; (2) the conservation of a few specific transcription factors, corresponding to multifunctional enzymes, that may have acquired a broader regulatory role; (3) the conservation of the regulator DksA, in the absence of the alarmone ppGpp, possibly to enhance transcription elongation in these AT-rich genomes and perhaps to discriminate between stable and unstable promoters, as suggested by Srivatsan and Wang in *E. coli* [102]; (4) the conservation of several NAPs and of their positioning organisation along the chromosome; (5) the conservation of unstable promoters upstream of transport and regulatory genes.

This work has allowed us to propose some assumptions about the role of DNA topology in gene expression in *Buchnera*. Hence, AT-bias is generally considered to be either a consequence of the degeneration of the repair system in *Buchnera* [103] or for energetic selection linked to the centrality of ATP in the metabolism of the bacterium [104]. We have proposed here that the AT-richness of the DNA molecule of *Buchnera* might provide a selective advantage, giving more flexibility and curvature to the chromosome and facilitating its regulation with only a small number of global regulators, such as NAPs and topoisomerases.

Finally, the presence of unstable promoters upstream of transporter genes might allow a basal expression of the transporter genes in *Bap*, even when conditions are unfavourable (potentially for importing metabolic precursors), and an over-expression of these genes when the energetic level of the cell is high (potentially to export metabolites for the host). Indeed, the ATP/ADP ratio, partly controlling the gyrase activity, is linked to the superhelicity of the DNA molecule in that when the ratio is high the level of DNA-supercoiling is high and DNA strands are destabilized by the constraints increasing the activity of most promoters. On the contrary, when the ratio is low the DNA molecule is more relaxed and only unstable promoters would be active.

Experimental studies will be needed to test whether such ancestral gene expression regulation by DNA-topology remains functional in *Bap* and in other *Buchnera* with even smaller genomes and could be a preponderant mechanism of gene transcription regulation in these bacterial cells with a very small diversity of transcription factors. Hence, DNA-topology could regulate some important symbiotic functions, as well as whether superhelical changes may have occurred during the few contrasted

physiological states imposed on *Buchnera* by their integration into the aphid life cycle (e.g., the establishment of symbiosis during embryonic development, the growing phase and activation of symbiotic metabolism, decaying symbiosis in old aphids, etc.).

Additional files

Additional file 1: (Figure): Functional analysis (GO terms) of *BAP* genes.

Additional file 2: (Figure): Conservation of the protein sequences of the six NAP encoding genes found in *Buchnera*.

Additional file 3: (Table): Number (#) of specific NAP target genes in *E. coli* (as found in the RegulonDB) and of the orthologous genes (putative conserved targets) in the four *Buchnera* strains *BAP*, *BSg*, *BBp* and *BCc*.

Additional file 4: (Table): Number (#) of predicted σ^{70} promoters in the four strains of *Buchnera*.

Additional file 5: (Figure): Distribution of σ^{70} prediction scores and 5'UTR lengths in *Bap* and *E. coli*.

Additional file 6: (Figure): Venn-diagram of the putative σ^{32} regulons in *Buchnera*.

Additional file 7: (Table): Transporter (T) and regulatory (R) genes of *BAP* significantly correlated with low SIDD values (i.e., associated with unstable promoter regions).

Abbreviations

BAP, *BBp*, *BCc*, *BCt*, *BSg*: *Buchnera aphidicola* from the species *Acyrtosiphon pisum*, *Baizongia pistacea*, *Cinara cedri*, *Cinara tujafilina* and *Schizaphis graminum*; CDS: Coding DNA Sequence; GO: Gene ontology; HTH: Helix-turn-helix; NAP: Nucleoid associated protein; SIDD: Stress-induced duplex destabilization; TFBS: Transcription factor binding site; TU: Transcription unit; UTR: Untranslated region.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

LB, FC and HC conceived and designed the study. LB performed the analyses. LB, FC and HC analyzed the data, contributed to the text, Tables and Figures, and wrote the paper. All authors read and approved the final version of the manuscript.

Acknowledgments

We would like to thank Sylvie Reverchon for her critical reading of this paper and Valerie James for the English corrections. This work was partly funded by the Agence National de la Recherche (ANR, France) and the Biotechnology and Biological Sciences Research Council (BBSRC, UK), within the MetNet4SyBio project. LB was the recipient of a PhD fellowship from the French Ministry of Research (2007–2010).

Author details

¹UMR203 BF2I, Biologie Fonctionnelle Insectes et Interactions, INSA-Lyon, INRA, Université de Lyon, Villeurbanne, France. ²BAMBOO, INRIA Rhône-Alpes, Montbonnot Saint-Martin, France. ³Present address: Université de Lyon, Université Lyon1, CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, Lyon, France.

Received: 17 October 2012 Accepted: 23 January 2013

Published: 1 February 2013

References

- Shigenobu S, Watanabe H, Hattori M, Sasaki Y, Ishikawa H: **Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* 2000, **407**(6800):81–86.**

2. Brinza L, Viñuelas J, Cottret L, Calevro F, Rahbé Y, Febvay G, Duport G, Colella S, Rabatel A, Gautier C, et al: **Systemic analysis of the symbiotic function of *Buchnera aphidicola*, the primary endosymbiont of the pea aphid *Acyrtosiphon pisum*.** *C R Biol* 2009, **332**(11):1034–1049.
3. Tamas I, Klasson L, Canback B, Naslund AK, Eriksson AS, Wernegreen JJ, Sandstrom JP, Moran NA, Andersson SGE: **50 million years of genomic stasis in endosymbiotic bacteria.** *Science* 2002, **296**(5577):2376–2379.
4. Van Ham RC, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, Fernandez JM, Jimenez L, Postigo M, Silva FJ, et al: **Reductive genome evolution in *Buchnera aphidicola*.** *Proc Natl Acad Sci U S A* 2003, **100**(2):581–586.
5. Pérez-Brocá V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, Silva FJ, Moya A, Latorre A: **A small microbial genome: the end of a long symbiotic relationship?** *Science* 2006, **314**(5797):312–313.
6. Lamelas A, Gosalbes MJ, Moya A, Latorre A: **New clues about the evolutionary history of metabolic losses in bacterial endosymbionts, provided by the genome of *Buchnera aphidicola* from the aphid *Cinara tujaefilina*.** *Appl Environ Microbiol* 2011, **77**(13):4446–4454.
7. Douglas AE: **The microbial dimension in insect nutritional ecology.** *Func Ecol* 2009, **23**(1):38–47.
8. Wilson AC, Ashton PD, Calevro F, Charles H, Colella S, Febvay G, Jander G, Kushlan PF, Macdonald SJ, Schwartz JF, et al: **Genomic insight into the amino acid relations of the pea aphid, *Acyrtosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*.** *Insect Mol Biol* 2010, **19**(Suppl 2):249–258.
9. Charles H, Balmard S, Lamelas A, Cottret L, Pérez-Brocá V, Burdin B, Latorre A, Febvay G, Colella S, Calevro F, et al: **A genomic reappraisal of symbiotic function in the aphid / *Buchnera* symbiosis: reduced transporter sets and variable membrane organisations.** *PLoS One* 2011, **6**(12):e29096.
10. Gosalbes MJ, Lamelas A, Moya A, Latorre A: **The striking case of tryptophan provision in the cedar aphid *Cinara cedri*.** *J Bacteriol* 2008, **190**(17):6026–6029.
11. Lamelas A, Gosalbes MJ, Manzano-Marin A, Pereto J, Moya A, Latorre A: ***Serratia symbiotica* from the aphid *Cinara cedri*: a missing link from facultative to obligate insect endosymbiont.** *PLoS Genet* 2011, **7**(11):e1002357.
12. Wilcox JL, Dunbar HE, Wolfinger RD, Moran NA: **Consequences of reductive evolution for gene expression in an obligate endosymbiont.** *Mol Microbiol* 2003, **48**(6):1491–1500.
13. Moran NA, Dunbar HE, Wilcox JL: **Regulation of transcription in a reduced bacterial genome: nutrient-provisioning genes of the obligate symbiont *Buchnera aphidicola*.** *J Bacteriol* 2005, **187**(12):4229–4237.
14. Reymond N, Calevro F, Viñuelas J, Morin N, Rahbé Y, Febvay G, Laugier C, Douglas AE, Fayard JM, Charles H: **Different levels of transcriptional regulation due to trophic constraints in the reduced genome of *Buchnera aphidicola* APS.** *Appl Environ Microbiol* 2006, **72**(12):7760–7766.
15. Bermingham J, Rabatel A, Calevro F, Viñuelas J, Febvay G, Charles H, Douglas A, Wilkinson T: **Impact of host developmental age on the transcriptome of the symbiotic bacterium *Buchnera aphidicola* in the pea aphid (*Acyrtosiphon pisum*).** *Appl Environ Microbiol* 2009, **75**(22):7294–7297.
16. Viñuelas J, Febvay G, Duport G, Colella S, Fayard JM, Charles H, Rahbé Y, Calevro F: **Multimodal dynamic response of the *Buchnera aphidicola* pLeu plasmid to variations in leucine demand of its host, the pea aphid *Acyrtosiphon pisum*.** *Mol Microbiol* 2011, **81**(5):1271–1285.
17. Marr C, Geertz M, Hutt MT, Muskhelishvili G: **Dissecting the logical types of network control in gene expression profiles.** *BMC Syst Biol* 2008, **2**:18.
18. Browning DF, Busby SJ: **The regulation of bacterial transcription initiation.** *Nat Rev Microbiol* 2004, **2**(1):57–65.
19. Blot N, Mavathur R, Geertz M, Travers A, Muskhelishvili G: **Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome.** *EMBO Rep* 2006, **7**(7):710–715.
20. Dorman CJ: **DNA supercoiling and environmental regulation of gene expression in pathogenic bacteria.** *Infect Immun* 1991, **59**(3):745–749.
21. Dorman CJ: **DNA supercoiling and bacterial gene expression.** *Sci Prog* 2006, **89**(Pt 3–4):151–166.
22. Ishihama A: **Prokaryotic genome regulation: multifactor promoters, multitarget regulators and hierarchic networks.** *FEMS Microbiol Rev* 2010, **34**(5):628–645.
23. Peter BJ, Arsuaga J, Breier AM, Khodursky AB, Brown PO, Cozzarelli NR: **Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*.** *Genome Biol* 2004, **5**(11):R87.
24. Pedersen AG, Jensen LJ, Brunak S, Staerfeldt HH, Ussery DW: **A DNA structural atlas for *Escherichia coli*.** *J Mol Biol* 2000, **299**(4):907–930.
25. Hallin P, Staerfeldt H, Rotenberg E, Binnewies T: **GeneWiz browser: an interactive tool for visualizing sequenced chromosomes.** *Stand Genomic Sci* 2009, **1**(2):204–215.
26. Brinza L, Calevro F, Duport G, Gaget K, Gautier C, Charles H: **Structure and dynamics of the operon map of *Buchnera aphidicola* sp. Strain APS.** *BMC Genomics* 2010, **11**:666.
27. Gottesman S: **Micros for microbes: non-coding regulatory RNAs in bacteria.** *Trends Genet* 2005, **21**(7):399–404.
28. Gripenland J, Netterling S, Loh E, Tiensuu T, Toledo-Arana A, Johansson J: **RNAs: regulators of bacterial virulence.** *Nature Rev Microbiol* 2010, **8**(12):857–866.
29. Thao S, Chen CS, Zhu H, Escalante-Semerena JC: **Nepsilon-lysine acetylation of a bacterial transcription factor inhibits its DNA-binding activity.** *PLoS One* 2010, **5**(12):e15123.
30. Murphy LD, Zimmerman SB: **Isolation and characterization of spermidine nucleoids from *Escherichia coli*.** *J Struct Biol* 1997, **119**(3):321–335.
31. Murphy LD, Zimmerman SB: **Stabilization of compact spermidine nucleoids from *Escherichia coli* under crowded conditions: implications for *in vivo* nucleoid structure.** *J Struct Biol* 1997, **119**(3):336–346.
32. Azam T, Ishihama A: **Twelve species of the nucleoid-associated protein from *Escherichia coli*.** *J Biol Chem* 1999, **274**:33105–33113.
33. Dillon SC, Dorman CJ: **Bacterial nucleoid-associated proteins, nucleoid structure and gene expression.** *Nat Rev Microbiol* 2010, **8**(3):185–195.
34. Dorman CJ: **H-NS: a universal regulator for a dynamic genome.** *Nat Rev Microbiol* 2004, **2**(5):391–400.
35. Travers A, Muskhelishvili G: **DNA supercoiling - a global transcriptional regulator for enterobacterial growth?** *Nat Rev Microbiol* 2005, **3**(2):157–169.
36. Zechiedrich EL, Khodursky AB, Bachellier S, Schneider R, Chen D, Lilley DM, Cozzarelli NR: **Roles of topoisomerases in maintaining steady-state DNA supercoiling in *Escherichia coli*.** *J Biol Chem* 2000, **275**(11):8103–8113.
37. Kreuzer KN, Cozzarelli NR: ***Escherichia coli* mutants thermosensitive for deoxyribonucleic acid gyrase subunit A: effects on deoxyribonucleic acid replication, transcription, and bacteriophage growth.** *J Bacteriol* 1979, **140**(2):424–435.
38. Cheung KJ, Badarinarayana V, Selinger DW, Janse D, Church GM: **A microarray-based antibiotic screen identifies a regulatory role for supercoiling in the osmotic stress response of *Escherichia coli*.** *Genome Res* 2003, **13**(2):206–215.
39. Sobetzko P, Travers A, Muskhelishvili G: **Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle.** *Proc Natl Acad Sci U S A* 2011, **109**(2):42–50.
40. Wang H, Benham CJ: **Superhelical destabilization in regulatory regions of stress response genes.** *PLoS Comput Biol* 2008, **4**(1):e17.
41. Ornstein R, Rein R, Breen D, Macelroy R: **An optimized potential function for the calculation of nucleic acid interaction energies I. Base stacking.** *Biopolymers* 1978, **17**(10):2341–2360.
42. Benham CJ, Bi C: **The analysis of stress-induced duplex destabilization in long genomic DNA sequences.** *J Comp Biol* 2004, **11**(4):519–543.
43. Price MN, Dehal PS, Arkin AP: **Orthologous transcription factors in bacteria have different functions and regulate different genes.** *PLoS Comput Biol* 2007, **3**(9):1739–1750.
44. Baumbach J, Rahmann S, Tauch A: **Reliable transfer of transcriptional gene regulatory networks between taxonomically related organisms.** *BMC Syst Biol* 2009, **3**:8.
45. Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, et al: **The complete genome sequence of *Escherichia coli* K-12.** *Science* 1997, **277**(5331):1453–1462.
46. Riley M, Abe T, Arnaud MB, Berlyn MK, Blattner FR, Chaudhuri RR, Glasner JD, Horiuchi T, Keseler IM, Kosuge T, et al: ***Escherichia coli* K-12: a cooperatively developed annotation snapshot - 2005.** *Nucleic Acids Res* 2006, **34**(1):1–9.
47. Gama-Castro S, Jimenez-Jacinto V, Peralta-Gil M, Santos-Zavaleta A, Penaloza-Spinola MI, Contreras-Moreira B, Segura-Salazar J, Muniz-Rascado L, Martinez-Flores I, Salgado H, et al: **RegulonDB (version 6.0): Gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and textpresso navigation.** *Nucleic Acids Res* 2008, **36**(Database issue):D120–D124.
48. Prickett MD, Page M, Douglas AE, Thomas GH: ***Buchnera*BASE: a post-genomic resource for *Buchnera* sp. APS.** *Bioinformatics* 2006, **22**(5):641–642.

49. Finn RD, Mistry J, Tate J, Coghill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, et al: **The Pfam protein families database.** *Nucleic Acids Res* 2010, **38**(Database issue):D211–D222.
50. Perez-Rueda E, Collado-Vides J, Segovia L: **Phylogenetic distribution of DNA-binding transcription factors in bacteria and archaea.** *Comp Biol Chem* 2004, **28**(5–6):341–350.
51. Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM: **The many faces of the helix-turn-helix domain: transcription regulation and beyond.** *FEMS Microbiol Rev* 2005, **29**(2):231–262.
52. Dodd IB, Egan JB: **Improved detection of helix-turn-helix DNA-binding motifs in protein sequences.** *Nucleic Acids Res* 1990, **18**(17):5019–5026.
53. Combet C, Blanchet C, Geourjon C, Deleage G: **NPS@: network protein sequence analysis.** *Trends Biochem Sci* 2000, **25**(3):147–150.
54. Wilson D, Charoensawan V, Kummerfeld SK, Teichmann SA: **DBD-taxonically broad transcription factor predictions: new content and functionality.** *Nucleic Acids Res* 2008, **36**(Database issue):D88–D92.
55. Gross CA: **Function and regulation of the heat shock proteins.** In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, Volume 1. Edited by Neidhardt FC. Washington, DC: American Society for Microbiology Press; 1996:1382–1399.
56. Charif D, Lobry J: **SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis.** In *Structural approaches to sequence evolution: Molecules, networks, populations*. Edited by Bastolla U, Porto M, Roman H, Vendruscolo M. New York: Springer Verlag; 2007:207–232.
57. Ussery D, Larsen TS, Wilkes KT, Friis C, Worning P, Krogh A, Brunak S: **Genome organisation and chromatin structure in *Escherichia coli*.** *Biochimie* 2001, **83**(2):201–212.
58. Goodrich JA, Schwartz ML, McClure WR: **Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for *Escherichia coli* integration host factor (IHF).** *Nucleic Acids Res* 1990, **18**(17):4993–5000.
59. Lang B, Blot N, Bouffartigues E, Buckle M, Geertz M, Gualerzi CO, Mavathur R, Muskhelishvili G, Pon CL, Rimsky S, et al: **High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes.** *Nucleic Acids Res* 2007, **35**(18):6330–6337.
60. Barrell D, Dimmer E, Huntley RP, Binns D, O'Donovan C, Apweiler R: **The GOA database in 2009 - an integrated gene ontology annotation resource.** *Nucleic Acids Res* 2009, **37**(Database issue):D396–D403.
61. Boyer F, Morgat A, Labarre L, Pothier J, Viari A: **Syntons, metabolons and interactons: an exact graph-theoretical approach for exploring neighbourhood between genomic and functional data.** *Bioinformatics* 2005, **21**(23):4209–4215.
62. Shpigelman ES, Trifonov EN, Bolshoy A: **CURVATURE: software for the analysis of curved DNA.** *Comput Appl Biosci* 1993, **9**(4):435–440.
63. Hassan M, Calladine C: **Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA.** *J Mol Biol* 1996, **259**:95–103.
64. Thomas GH, Zucker J, Macdonald SJ, Sorokin A, Goryanin I, Douglas AE: **A fragile metabolic network adapted for cooperation in the symbiotic bacterium *Buchnera aphidicola*.** *BMC Syst Biol* 2009, **3**:24.
65. Seshasayee ASN, Fraser GM, Babu MM, Luscombe NM: **Principles of transcriptional regulation and evolution of the metabolic system in *E. coli*.** *Genome Res* 2009, **19**(1):79–91.
66. Poliakov A, Russell CW, Ponnala L, Hoops HJ, Sun Q, Douglas AE, van Wijk KJ: **Large-scale label-free quantitative proteomics of the pea aphid-*Buchnera* symbiosis.** *Mol Cell Proteomics* 2011, **10**(6):M110.007039.
67. Moran NA, Mira A: **The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*.** *Genome Biol* 2001, **2**(12):RESEARCH0054.
68. Silva FJ, Latorre A, Moya A: **Genome size reduction through multiple events of gene disintegration in *Buchnera* APS.** *Trends Genet* 2001, **17**(11):615–618.
69. Putney SD, Schimmel P: **An aminoacyl tRNA synthetase binds to a specific DNA sequence and regulates its gene transcription.** *Nature* 1981, **291**(5817):632–635.
70. Charlier D, Kholti A, Huysveld N, Gigot D, Maes D, Thia-Toong TL, Glansdorff N: **Mutational analysis of *Escherichia coli* PepA, a multifunctional DNA-binding aminopeptidase.** *J Mol Biol* 2000, **302**(2):411–426.
71. Huynen MA, Spronk CA, Gabaldon T, Snel B: **Combining data from genomes, Y2H and 3D structure indicates that BolA is a reductase interacting with a glutaredoxin.** *FEBS Lett* 2005, **579**(3):591–596.
72. Yeung N, Gold B, Liu NL, Prathapam R, Sterling HJ, Williams ER, Butland G: **The *E. coli* monothiol glutaredoxin GrxD forms homodimeric and heterodimeric FeS cluster containing complexes.** *Biochemistry* 2011, **50**(41):8957–8969.
73. Santos JM, Freire P, Vicente M, Arraiano CM: **The stationary-phase morphogene *bolA* from *Escherichia coli* is induced by stress during early stages of growth.** *Mol Microbiol* 1999, **32**(4):789–798.
74. Aldea M, Garrido T, Hernandez-Chico C, Vicente M, Kushner SR: **Induction of a growth-phase-dependent promoter triggers transcription of *bolA*, an *Escherichia coli* morphogene.** *EMBO J* 1989, **8**(12):3923–3931.
75. Santos JM, Lobo M, Matos AP, De Pedro MA, Arraiano CM: **The gene *bolA* regulates *dacA* (PBP5), *dacC* (PBP6) and *ampC* (AmpC), promoting normal morphology in *Escherichia coli*.** *Mol Microbiol* 2002, **45**(6):1729–1740.
76. Freire P, Moreira RN, Arraiano CM: **BolA inhibits cell elongation and regulates MreB expression levels.** *J Mol Biol* 2009, **385**(5):1345–1351.
77. Bachler C, Schneider P, Bahler P, Lustig A, Erni B: ***Escherichia coli* dihydroxyacetone kinase controls gene expression by binding to transcription factor DhaR.** *EMBO J* 2005, **24**(2):283–293.
78. Shen A, Kamp HD, Grundling A, Higgins DE: **A bifunctional O-GlcNAc transferase governs flagellar motility through anti-repression.** *Genes Dev* 2006, **20**(23):3283–3295.
79. Commichau FM, Herzberg C, Tripal P, Valerius O, Stulke J: **A regulatory protein-protein interaction governs glutamate biosynthesis in *Bacillus subtilis*: the glutamate dehydrogenase RocG moonlights in controlling the transcription factor GltC.** *Mol Microbiol* 2007, **65**(3):642–654.
80. Hullo MF, Auger S, Soutourina O, Barzu O, Yvon M, Danchin A, Martin-Verstraete I: **Conversion of methionine to cysteine in *Bacillus subtilis* and its regulation.** *J Bacteriol* 2007, **189**(1):187–197.
81. Bae W, Xia B, Inouye M, Severinov K: ***Escherichia coli* CspA-family RNA chaperones are transcription antiterminators.** *Proc Natl Acad Sci U S A* 2000, **97**(14):7784–7789.
82. Phadtare S, Inouye M: **Sequence-selective interactions with RNA by CspB, CspC and CspE, members of the CspA family of *Escherichia coli*.** *Mol Microbiol* 1999, **33**(5):1004–1014.
83. Edwards AN, Patterson-Fortin LM, Vakulskas CA, Mercante JW, Potrykus K, Vinella D, Camacho MI, Fields JA, Thompson SA, Georgellis D, et al: **Circuitry linking the Csr and stringent response global regulatory systems.** *Mol Microbiol* 2011, **80**(6):1561–1580.
84. Romeo T: **Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB.** *Mol Microbiol* 1998, **29**(6):1321–1330.
85. Yakhnin H, Yakhnin AV, Baker CS, Sineva E, Berezin I, Romeo T, Babitzke P: **Complex regulation of the global regulatory gene *csrA*: CsrA-mediated translational repression, transcription from five promoters by σ and σ 54(S), and indirect transcriptional activation by CsrA.** *Mol Microbiol* 2011, **81**(3):689–704.
86. Perederina A, Svetlov V, Vassilyeva MN, Tahirov TH, Yokoyama S, Artsimovitch I, Vassilyev DG: **Regulation through the secondary channel - structural framework for ppGpp-DksA synergism during transcription.** *Cell* 2004, **118**(3):297–309.
87. Meddows TR, Savory AP, Grove JI, Moore T, Lloyd RG: **RecN protein and transcription factor DksA combine to promote faithful recombinational repair of DNA double-strand breaks.** *Mol Microbiol* 2005, **57**(1):97–110.
88. Haugen SP, Ross W, Manrique M, Gourse RL: **Fine structure of the promoter-sigma region 1.2 interaction.** *Proc Natl Acad Sci U S A* 2008, **105**(9):3292–3297.
89. Rutherford ST, Villers CL, Lee JH, Ross W, Gourse RL: **Allosteric control of *Escherichia coli* rRNA promoter complexes by DksA.** *Genes Dev* 2009, **23**(2):236–248.
90. Vinella D, Potrykus K, Murphy H, Cashel M: **Effects on growth by changes of the balance between GreA, GreB, and DksA suggest mutual competition and functional redundancy in *Escherichia coli*.** *J Bacteriol* 2012, **194**(2):261–273.
91. Li K, Jiang T, Yu B, Wang L, Gao C, Ma C, Xu P, Ma Y: **Transcription elongation factor GreA Has functional chaperone activity.** *PLoS One* 2012, **7**(12):e47521.
92. Guinote IB, Moreira RN, Freire P, Arraiano CM: **Characterization of the BolA homolog lbaG: a new gene involved in acid resistance.** *J Microbiol Biotechnol* 2012, **22**(4):484–493.
93. Viñuelas J, Calevro F, Remond D, Bernillon J, Rahbe Y, Febvay G, Fayard JM, Charles H: **Conservation of the links between gene transcription and chromosomal organization in the highly reduced genome of *Buchnera aphidicola*.** *BMC Genomics* 2007, **8**(1):143.
94. Kano Y, Imamoto F: **Requirement of integration host factor (IHF) for growth of *Escherichia coli* deficient in HU protein.** *Gene* 1990, **89**(1):133–137.

95. Komaki K, Ishikawa H: Intracellular bacterial symbionts of aphids possess many genomic copies per bacterium. *J Mol Evol* 1999, **48**(6):717–722.
96. Gil R, Silva FJ, Pereto J, Moya A: Determination of the core of a minimal bacterial gene set. *Microbiol Mol Biol Rev* 2004, **68**(3):518–537.
97. Drolet M, Phoenix P, Menzel R, Masse E, Liu LF, Crouch RJ: Overexpression of RNase H partially complements the growth defect of an *Escherichia coli* delta *topA* mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. *Proc Natl Acad Sci U S A* 1995, **92**(8):3526–3530.
98. Moran NA: Tracing the evolution of gene loss in obligate bacterial symbionts. *Curr Opin Microbiol* 2003, **6**(5):512–518.
99. Stoll S, Feldhaar H, Gross R: Promoter characterization in the AT-rich genome of the obligate endosymbiont "*Candidatus Blochmannia floridanus*". *J Bacteriol* 2009, **191**(11):3747–3751.
100. Riley M: Genes and proteins of *Escherichia coli* K-12. *Nucleic Acids Res* 1998, **26**(1):54.
101. Balleza E, Lopez-Bojorquez LN, Martinez-Antonio A, Resendis-Antonio O, Lozada-Chavez I, Balderas-Martinez YI, Encarnacion S, Collado-Vides J: Regulation by transcription factors in bacteria: beyond description. *FEMS Microbiol Rev* 2009, **33**(1):133–151.
102. Srivatsan A, Wang JD: Control of bacterial transcription, translation and replication by (p)ppGpp. *Curr Opin Microbiol* 2008, **11**(2):100–105.
103. Tamas I, Wernegreen JJ, Nystedt B, Kauppinen SN, Darby AC, Gomez-Valero L, Lundin D, Poole AM, Andersson SG: Endosymbiont gene functions impaired and rescued by polymerase infidelity at poly(A) tracts. *Proc Natl Acad Sci U S A* 2008, **105**(39):14934–14939.
104. Rocha EPC, Danchin A: Base composition bias might result from competition for metabolic resources. *Trends Genet* 2002, **18**(6):291–294.

doi:10.1186/1471-2164-14-73

Cite this article as: Brinza et al.: Genomic analysis of the regulatory elements and links with intrinsic DNA structural properties in the shrunken genome of *Buchnera*. *BMC Genomics* 2013 **14**:73.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

